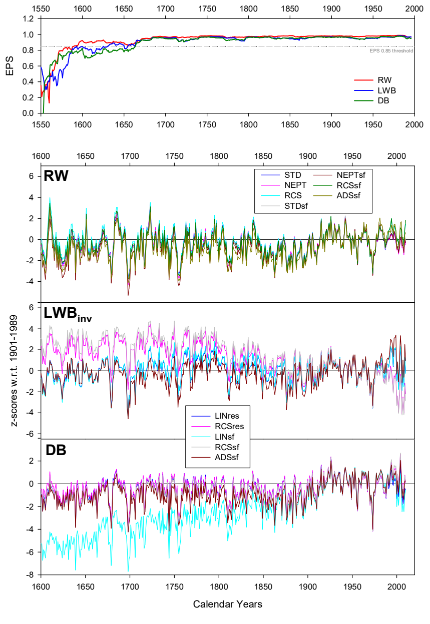
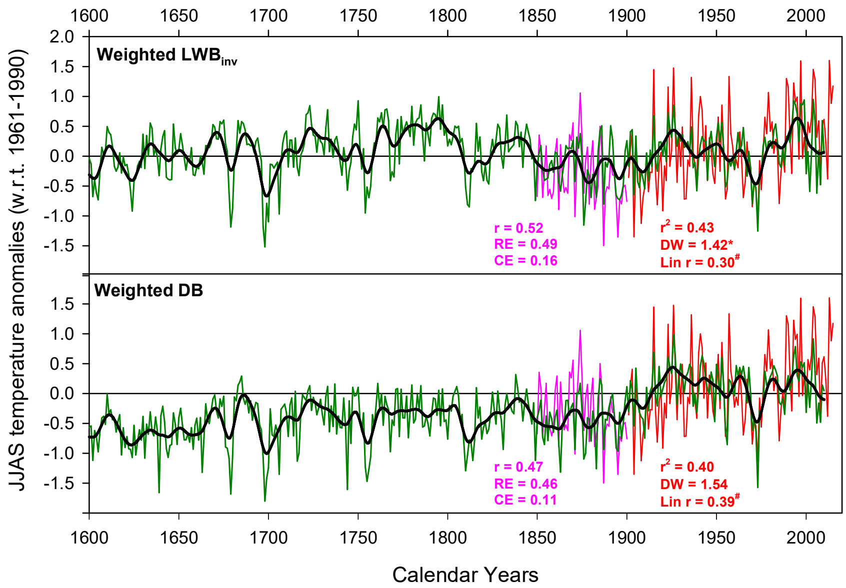
Review 1

General comments:

The major problem with this manuscript is the combination of a great number of models tested and the great variability among the different models tested. Because of this combination it could be argued that the successful models have been achieved spuriously. ***We think this is a little unfair as this is an exploratory paper which presents a range of results which provide guidance for the continuation of this work as we extend these new data back into the 1st millennium AD. We present a series of calibration experiments using each TR parameter separately and varying combinations of parameters. The main conclusion from this admittedly limited initial data-set is that the DB parameter likely expresses the best compromise between high and low frequency calibration fidelity. We find it strange that the reviewer thinks that we have attained a spurious result as we have been cautious with our conclusions.***

We are however willing to make some changes to allay the reviewer's concerns. An age dependent spline chronology version can be added to Figure 6. For Figure 7, rather than use a single reconstruction, we can derive a LWB and DB based GOA composite reconstruction based on all the chronology variants as weighted means related to their r2 values to the 1901-2010 calibration period. The new Figures 6 and 7 are shown below and the text can be changed appropriately.





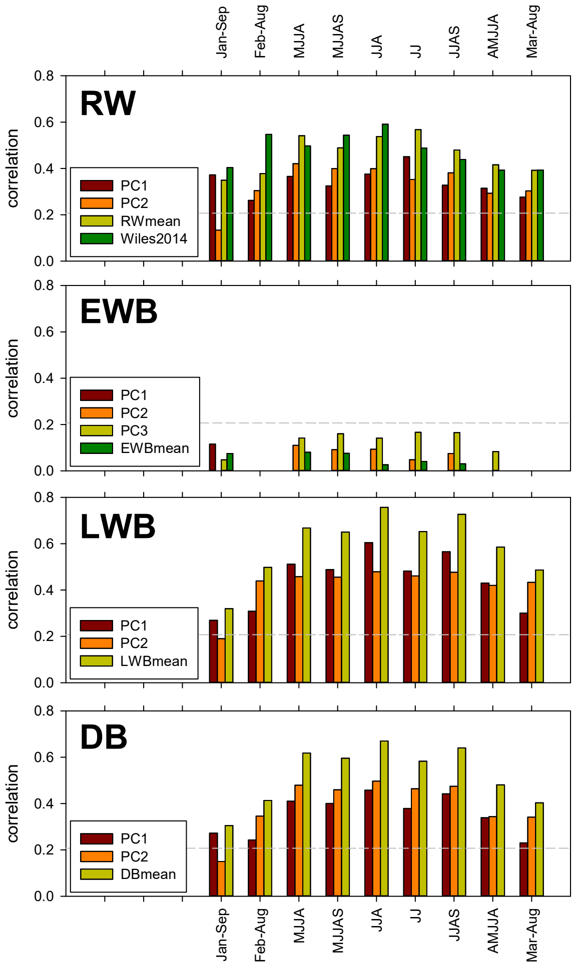
To attempt to avoid this, I recommend to limit the number of models tested, by performing also climate calibrations with high-pass filtered data (see suggestions below) and to combine this with a discussion of which monthly temperatures can have a causal effect on tree growth. Conducting this additional analysis would narrow down the options that can be tested, but also function as a baseline for the discussion of low-frequency skill in the data. If the high-frequency part of the data is agreeing well with temperature, it is likely safe to assume that a breakdown of agreement when low-frequencies are added is due to low-frequency biases, such as HW-SW-, standardization, etc.. problems. When this is established then tests of how to minimize the loss of signal at lower frequencies can be conducted (different standardizations alternatives). If however, the high-frequency part of the data does not agree very well with temperatures in the first place, it is very unlikely to expect that adding the low-frequency part will contribute with useful information even if correlations are boosted. Therefore, the high-frequency analysis must come first and inform subsequent choices of configurations and options.

Dendroclimatology, unlike most other palaeoclimate approaches has the ability to derive so-called robust estimates of past climate at inter-annual to centennial time-scales. Although we agree that low frequency trends can lead to spurious correlations, we have been very careful in testing both the high and low frequency fidelity of the models we present. The low frequency fidelity is particularly difficult to examine as we must assume greater uncertainty in the early instrumental data. In fact, the ambiguity of this paper is the assessment of the low frequency signal in the LWB and DB data.

However, the reviewer is entirely correct that we ideally want "equal" coherence with past temperatures at both high and low frequencies. Rather than add further calibration experiments using more flexible detrending options, we feel that a valid compromise is to also present (as an extra supplementary figure) the correlation response function analysis results of Figure 2 after the data (TR and temperature) have been transformed to 1st differences. This new figure version is presented below.

The RW based correlations show strongest correlations with JJA, but are still relatively high with the broader window of Feb-Aug. The reviewer is therefore correct that there is some spurious trend related correlation creeping in for RW, but is only minor and arguably irrelevant for this paper as the focus of the paper is on the Blue Intensity based parameters anyway.

The EWB 1st differenced correlations at inter-annual time-scales are non-significant. For LWB and DB, the correlations are overall similar as the original figure 2 although the seasons including winter months are weaker. Again – this does not impact the paper as the summer months were the target calibration seasons. Again, as with the original figure 2, the correlations are generally higher for LWB than DB and ultimately, the low frequency issue (i.e. potential biased in LWB) become much more significant. This result has not changed. The main issue is which of JJA or JJAS are the optimal season to calibrate against and how one combines the different parameters – i.e. use RW and DB in a multiple regression, or utilise a band-pass approach and use the RW and LWB data at the frequencies where their signal is "robust" – see Rydval et al. 2017 for an example.



A secondary issue is that the authors use reflected BI. This type of data is negatively correlated with what the authors claim to measure in the wood: cell wall, lignin content, but also with the discolorations. If the authors would opt to use the absorbed BI it would let them completely avoid many confused elaborations (see detailed comments below) with regard to standardization and comparisons with MXD etc.

This is a semantic comment about terminology and does not impact the analysis in any way. I believe that our methodological description is clear although a minor clarification is possible (see below).

As lignin absorbs [blue] light, then dense latewood (high density) will reflect less blue light. Hence raw LWB and MXD are inversely correlated. The only reason that raw LWB need to be inverted is due to limitations of the freely available detrending software (i.e. Arstan) where it is the norm to remove negative/zero slope trends and retain positive trends. Theoretically, the trends in LWB will be opposite to MXD, but in the software there are no options to remove positive/zero slope trends and retain negative ones. Hence the raw LWB data need to be inverted for detrending. This approach has been used in Wilson et al. (2011, 2014, 2017) and Rydval et al (2014, 2016, 2017) plus other papers and as far as we are aware it is only Björklund et al. who have suggested using the term "maximum latewood blue absorption intensity (MXBI)".

As a subtle re-wording we are willing to tweak the methodological text as follows:

"Raw EWB and LWB variables were measured using CooRecorder 8.1 software (Cybis 2016 - http://www.cybis.se/forfun/dendro/index.htm), which has state-of-the-art capabilities to acquire accurate reflectance intensity RGB colour measurements from scanned wood samples (see Rydval et al. 2014). DB values were calculated within CooRecorder by subtracting the raw LWB values from the raw EWB values for each year. Since raw LWB is negatively correlated to MXD (high density 'dark' latewood = low reflectance), values were inverted following the method detailed in Rydval et al. (2014) to allow for LWB (hereafter denoted as LWB_{inv}) to be detrended in a similar way to MXD (see also Wilson et al. 2014). The nature of the DB calculation results in this parameter being positively correlated with inverted LWB_{inv}, so these data could also be theoretically detrended in a similar way."

As a further compromise to the reviewer, we are also happy to add in a clear statement referring to Björklund et al. (2014/2015) stating their terminology for inverted LWB.

Finally, we believe it is important to treat BI related parameters independently of density. They are related no doubt, but trying to fit BI to density is a potentially dangerous approach. Although measuring similar properties, we cannot expect them to be exactly similar – especially w.r.t. age related trends. Also - it is not really clear what maximum early wood reflectance (EWB) actually represents and it is likely that this parameter is related to lumen size rather than any property reflecting compounds in the earlywood cell walls.

In conclusion, I find the manuscript well written and prepared but I strongly suggest adding a high-frequency analysis, and using absorbed BI. After these revisions and the implementation of the comments below, the manuscript should be suitable for publication. ***We hope the 1^{st} differenced based results shown above can address the reviewer's concerns on the first point and we feel no obligation to change our terminology to address the second point as it does not change the analysis/results in any way and we feel our current description is adequate and consistent with most previous publications.***

Detailed comments:

L45 Remove “However”
Agreed – this can be done when “allowed” to edit the paper for final publication

L49 replace “for” with “covering”
Agreed – this can be done when “allowed” to edit the paper for final publication

L54-59 This section only discuss the non-climatological variance distorting RW signal, and does not acknowledging that RW and LWB actually may contain different climatic fingerprints. I suggest adding something along this line.
Please see 1st differenced correlation response function analysis results above. These can be added to the supplementary and appropriate discussion added.

L70 Björklund et al worked with absorbed Maximum BI
Björklund’s absorbed Maximum BI is the same as the inverted LWB used in this paper. We feel the current description is clear enough – esp. with suggestions mentioned above.

L71-72 Here and in many other places it would be much simpler to start talking about absorbed BI values because these values will be positively correlated with the properties that you mention as potential measurement targets. Why measure the inverted value of what you are interested in? In this way just confuse readers about what you had to do before standardization to make them work properly and why BI is inversely correlated with density etc.
See comments above. I am sure Björklund measured raw intensity values of blue reflectance and then inverted the data as we have done. The difference after that is semantic only.

L80-81 This is true if we disregard the principle of diminishing records back in time.
This holds true despite the reduction in replication back in time. The community needs to be careful as to define clear threshold of truncation. In this paper, the data are far from ideal w.r.t. replication – that was partly strategic. The fact that the results are encouraging suggests that calibration will improve substantially as replication is increased. Not sure this comment warrants any specific change.

L92-93 Björklund et al 2014 subtracted average absorption Earlywood BI from maximum absorption BI.
This again highlights why we prefer our current methodological description. We have used the raw EWB and LWB values and used the difference to derive the DB value. This is mathematically the same as what the reviewer wants us to implement, but we see no gain with such a change as we are doing the same method, but using different terminology which is clearly defined.

L96-98 This sounds like a hypothesis you are going to test later in this paper, but it is not really tested. I would phrase it more like a discussion point: If EWB and LWB contain similar climatic responses and similar standard deviations. . .
Agreed – this is a little vague and reviewer 2 also flagged this. We would gladly change this text (and associated later discussion) to basically highlight that if EWB and LWB both express the same climatic response (this should not be the case), the resultant derived DB data will not show this common response and likely be inferior to LWB. This is an evolving property of DB and certainly needs more examination using more species and locations.

L100-101 Not really "another concern". I suggest changes to something like this (I let you worry bout the grammar and English): Finally, although BI based variables hold great promise as an alternative proxy to MXD at inter-annual time-scales, the potential ability of BI to capture decadal to centennial time-scales related to long term-climate changes is still under question.
Agreed – happy to re-write this section and clarify the message better.

L102 Please clarify if you mean HW-SW color difference
Agreed. Yes – HW-SW – this can be clarified when "allowed" to edit the paper for final publication

L131-134 If you decide to use absorbed BI values this entire section can be removed. If you decide to keep it as is, I strongly recommend to go in to a discussion about why the detrending alternatives are sensitive to this. For example, deterministic detrending such as Neg. exp. or hugershof assume a decline in data values with age. If data values instead have an assumed increase, these methods will be useless. The reason for wanting this added discussion is that some researchers have missed this point and use these methods also for reflected BI.
As discussed above, we are happy with our current terminology and methods description.

L136-138 I recommend expanding this to also include a more aggressive detrending, perhaps a 25- 35-year spline. This will give more robust climate correlation result. If there are lingering trends in the tree-ring data, and there will be some using 200-year-splines, the risk of spurious trend correlations is relatively high. Adding a high frequency alternative can help to better identify important months for tree-growth. I suspect that some months enter your models just because they have similar trends as the tree-ring data. Also, before performing the aggressive alternative, the climate data should also be detrended similarly. Furthermore, I recommend to restructure the presented results; The high frequency monthly data analysis should be in the main manuscript and the seasonal climate correlations in the supplement together with the low-frequency counterparts. The HW-SW problem will still be present in the analysis using a 200-year spline, if you want to remove this for the analysis you need a softer spline. Rbar, PCA, climate response, between variable correlations should all be done with data with less autocorrelation: softer spline. The low-frequency alternative can be presented on the background of this analysis, but not stand alone. The models' monthly targets for reconstruction should be informed in the first

place with high-frequency results. A discussion can be conducted referring to the low-frequency results but not as a major informant of the models.
The mean/median segment lengths of all sites is > 200 years so any "lingering" trends for individual series would be minimal. We can add a comment to this end.

We hope the added 1st differenced based CRFA will address the reviewer's concerns as to our analysis and the identification of the "correct" month for calibration.

The reviewer is suggesting a major re-analysis here and it is not clear what gain there would be. The 200-yr spline approach is a compromise between minimising the HW-SW bias while retain some realistic multi-decadal information. We see no justification of re-doing the analysis using a much more flexible spline.

L167 Please specify which function was used to model the regional curve.
Yes – sorry – we can add this information in. The regional curve was smoothed with a spline of 10% the RC series length and that function used for detrending.

L171-172 LINres has been shown to create quite some bias in resulting chronologies, see works of Melvin and Briffa, especially if used to model the RC. I instead recommend time-varying response smoother Melvin et al., 2007.
Agreed that a LINres approach MAY impart biases, but we would argue that all detrending approaches have their own biases. That is why we have presented a small sub-set of possible detrending choices. We could have expanded on this substantially, but that will be a focus when we have the full 1000+ year data-set. I would likely to remind the reviewer that the LINres approach has been used on many studies in the past.

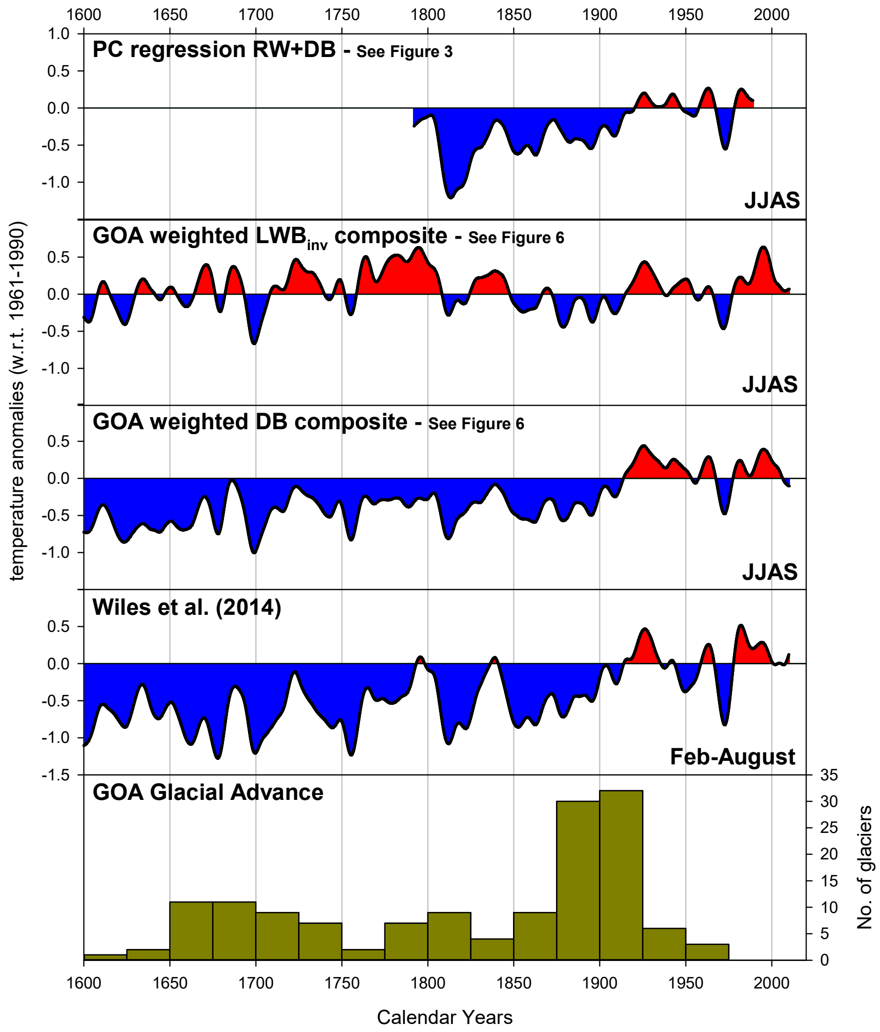
More importantly, the age-dependent spline approach of Melvin et al. (2007) is very much an untested detrending approach. I have experimented with this option and (1) can often inflate recent period values and (2) implicitly will remove secular scale variation – it is still a spline approach.

However, we are happy to add in an age dependent spline version into the mix for Figures 6 and 7 – I have shown these updates above. Table 5 can be updated:

	1901-2010 Calibration					1850-1900 Validation		
	series entered	r	r2	DW	LINr	r	RE	CE
LWB	LINres	0.64	0.41	1.36	0.36	0.53	0.44	0.07
	RCSres	0.26	0.07	1.28	0.48	0.56	0.01	-0.64
	LINsf	0.64	0.41	1.37	0.36	0.53	0.43	0.06
	RCSsf	0.21	0.05	1.32	0.46	0.56	-0.05	-0.73
	ADSsf	0.69	0.47	1.58	0.06	0.50	0.51	0.20

	1901-2010 Calibration					1850-1900 Validation		
	series entered	r	r2	DW	LINr	r	RE	CE
DB	LINres	0.55	0.31	1.37	0.50	0.50	0.52	0.21
	RCSres	0.64	0.40	1.59	0.40	0.48	0.50	0.18
	LINSF	0.54	0.29	1.35	0.38	0.43	0.40	0.00
	RCSsf	0.65	0.43	1.64	0.35	0.47	0.48	0.15
	ADSsf	0.65	0.42	1.59	0.30	0.47	0.34	-0.09

The r2 values for the 1901-2010 period can then be used as weighted in combining all these different variants to derive GOA weighted composites for parameter. The updated Figure 7 was shown earlier. The updated Figure 8 (which now includes the weighted LWB data) is shown below, which clearly shows that despite reasonable calibration and verification (Figure 7, table 5), the low frequency trends of the LWB data do appear at odds with the other data-sets.



L177-181 I suspect the results could be somewhat different with the high-frequency data analysis, see recommendations above. If they are, this is going to be vital information for your main question in the introduction: b) whether meaningful low frequency information can be gleaned from these data? Furthermore, if they are very different, the continuation of the question: "exploiting the long monthly instrumental temperature records that go back into the mid-19 validate secular trends in the TR data" becomes heavily diluted.

The current analysis, as a first attempt, already address the low frequency issue with appropriate discussion and the 1st difference results validate well that the appropriate season has been targeted.

Ultimately, it will never be possible to identify the "correct" detrending approach and when the final data have been finalised we will approach the reconstruction using a similar ensemble approach as introduced by Wilson et al. (2014) for a similar study in the Canadian Rockies. This will allow detrending uncertainty to be evaluated in the error estimates. At this time, we just wanted to highlight the sensitivity to such methodological choices – Figure 6 does this well.

L198 Again, must be done also with high-frequency data. Should likely cut off some month, and give a better causal reflection of which months are important for radial tree growth. High and persistent correlations with consecutive months makes me suspect trend-correlations.
See above. The 1st differenced CRFA addresses this.

L217-219 Awkward sentence, please rephrase.
This sentence can be tweaked when "allowed" to edit the paper for final publication

L227 in both or just in the new one?
We believe the current wording is quite clear that this is only for the latter "new data" RW based recon.

L265-271 Use absorption BI to avoid confusing comparisons with MXD.
See comments above.

L272-273 The original DB was introduced in Björklund et al., (2014), but it was further developed in Björklund et al., (2015) were they used a contrast adjustment. More discolored samples had a systematically lower contrast between earlywood and latewood than less discolored samples. If there is a systematic difference in discoloration then this will affect also the traditional DB data. You can easily test if there is a contrast problem in your data with scatterplots of DB vs EWBI, as done in Björklund et al., (2015). If there is a relationship you might at least want to discuss this. If there is not a relationship you will have cleared a question mark.
This is a good point, but is not the relevant analysis to be performed for this current paper. We believe for the current data adaptive detrending techniques used herein, that this issue is not yet relevant, BUT, will become relevant as we incorporate snag and sub-fossil material to extend the regional composite chronology back in time.

L276-281 According to my experience the age-trend of MXD would be more similar to DB than LWB. Perhaps different detrending options are needed, but if age-dependent splines are used, as suggested before, these would adapt to the small differences in the data. Neg. exp. or linear functions, for instance, may be directly inappropriate when having juvenile phases of increase and then followed by a decline.

As previously mentioned the current "conservative" approaches aims to retain low frequency information. We believe the age dependent spline will remove such trends. However, an age dependent spline option is now included in the analysis.

L278 Again use absorption BI.
See above.

L283-288 Use absorption BI to avoid having to clarify what you mean.
No – see above. The discussion will not change by using different terminology.
L288-292 It seems as a contradiction to write that LWB (as temperature proxy) should not have a negative trend w.r.t glacier advancements? The glacier advancement was stable up until 1800 CE and glacier advancement peaked around the turn of the 20th century. Would fit very well with the LWB record that has no trend from 1600-1800 CE and then a negative trend from 1800-1900 CE. The problem would be that there is no pronounced positive trend in the 20th century to melt away the glaciers that expanded prior to this.
We believe the reviewer is perhaps a little confused here. All LWB variants imply warmer than average (20th century) temperatures prior to 1850. This does NOT fit with the glacial expansion data shown in Figure 8. The DB based reconstruction (and variants) however, denote cooler than average (20th century) temperatures prior to 1850 which is in line with cooler conditions needed for continued glacial expansion through this period.

L343 Conclusions sections is very long and more like a summary of the discussion
1.5 pages of summary and recommendations appear appropriate to us. In fact, a little more discussion may be added about potential future strategies to address the high and low frequency issues of the LWB vs DB vs RW data.

L349-353 I would recommend to test high frequency results before making these bold statements. That is, to first to rule out any trend correlations with winter months for ring-width. After all it is very unusual for ring width to have a broader temperature response than Bl or density se e.g. Briffa et al., (2002).
Beyond the scope of this paper, if the RW data are regressed on the DB data, the residuals from this analysis correlate with winter temperatures. See Wilson et al. (2007) where the RW composite clearly picks up decade scale shifts seen in the PDO and other metrics of Pacific Decadal Variability. Such shifts are NOT seen in the BI based parameters.

Review 2

Specific comments:

Considering the experimental nature of the LWB and particularly DB parameters, it would have been useful to develop even a limited MXD dataset on at least part of the samples (e.g. from one site) in order to enable a direct comparison of the lower frequency trends in the BI data. Although it is argued that the structure of mountain hemlock wood makes it more difficult to prepare and measure these samples for density, it is not impossible and has been done in past studies. This would have been helpful in evaluating and constraining the utility of differently detrended BI chronologies and therefore considerably benefited this study in further strengthening the case for the use of DB as a better, less biased parameter relative to LWB and a suitable alternative to MXD for this species, especially since this is the first study to measure BI on mountain hemlock samples. Was this option at all considered?

MXD has not been measured on mountain Hemlock trees in this region as far as we are aware. Although we agree theoretically with the reviewer's comment that this comparison could be useful, MXD was never factored into the research design or budget and the focus of the study was always on BI and related parameters. We should also be careful not to assume that MXD is the "truth".

I am somewhat surprised that a higher number of samples was not used for the individual sites. According to Table 1 replication should ideally be 12-28 series for LWB and 14-36 for DB depending on the site. In several cases, the actual maximum number of series used is below (and in some cases well below) this optimal level. Is this not a problem? The weaker signal strength of BI data and the need for higher replication in order to develop 'robust' chronologies is acknowledged (e.g. L368-369). Also, the relatively low replication may even affect the RW data as stated in L229-230. As stated in the text, a subset of samples was selected for this study from earlier work so why not aim for 25-30 samples per site? That would have at least reduced uncertainty about the representativeness of some of the BI site chronologies, particularly in earlier periods when replication is likely even lower. It would be nice to see a replication plot over time (and EPS plot) for separate sites as well as for the 'all series' pooled version (perhaps as an SI figure) to give a better indication of which periods might potentially be affected by low replication.

This is an explorative paper and we have been clear, as the reviewer states that the number of records are lower than would be ideal. However, as we continue to develop these records (data being added continually) we wanted to write this initial explorative paper to partly inform ourselves as to the continued strategy of this work as well as the wider community. Previous RW based calibrations (Wilson et al 2007; Wiles et al. 2014) show that when replication is high, we can explain around 40% of the temperature variance. Using this non-ideal data-set we only explain 27% for RW, but >40% using LWB or DB. This suggests that the results can only improve as we increase replication. We are happy to make a stronger specific comment in this regard.

L96-98 and L217-219 – What is the rationale for this statement? To my knowledge this issue has not been investigated in any previous study. Presumably a higher correlation between EWB and LWB would imply that EWB expresses in part the same information as

LWB, but does that necessarily mean that this information is related to climate? How do you define 'weakly correlated'? Or in other words, what correlation would be acceptable and what would not? Ideally, this statement could be supported with an example and actual data. If nothing else, I would suggest elaborating further on this statement to more clearly express the justification for this claim.
This was brought up by Review 1 and we agree the current wording is ambiguous and needs clarification. Basically we want to provide a theoretical basis that DB will likely be useful when EWB and LWB are uncorrelated at high frequencies.

L102-105 As a general comment, some of the limitations of BI (specifically LWB) have already been explored in other studies. Clearly DB is a major improvement, although I wonder just how well DB resolves these issues and specifically whether DB could still have some problems at lower or other frequencies. It is interesting that in some cases the calculation of DB weakens the common signal, suggesting that information which should ideally be preserved is to some extent being removed in the process, yet in other instances the strength of the common signal remains relatively unaffected or even shows improvement. I suppose these questions can only be answered by various future studies that will further explore DB and I would not expect this to be fully covered here. But perhaps a statement could be included somewhere to caution and emphasize that considerable uncertainty remains with respect to the performance of DB and more work is required in this area.
This is a valid comment and we should emphasise better that in fact the year-to-year signal in the LWB data is often stronger than the DB data. Although not relevant for this specific paper, a similar band-pass approach as utilised by Rydval et al. (2016 and 2017) could work very well in the GOA region. We can add more discussion in this regard but this will be the focus for the final paper.

L109-111 - Is there any indication to what degree early instrumental biases could be a limitation (if at all) in achieving the stated aim?
This is an important point and we are happy to add further discussion about the issues of using early instrumental data.

(L133-138) Is there actually any need to detrend DB series? What is the justification for this? Hypothetically, if both LWB and EWB contain the same ontogenetic trends then by the nature of the DB calculation this trend would be automatically removed. I do not know whether or not that is true. This may be a more complex issue - perhaps only the LWB contains this trend or the LWB and EWB trends related to age differ in some way. But is it not possible that by detrending the DB data some lower frequency climatic information may be unnecessarily removed? Was the development of DB chronologies without performing detrending considered or explored in the analysis? The DB chronology in Figure 5 actually looks like a reasonable chronology variant and so I wonder how non-detrended DB chronologies would perform in terms of calibration and validation statistics relative to the detrended versions.
This is a valid point. Theoretically, one would assume that DB would be similar to MXD in trend, but I don't think the experimentation with BI based parameters are advanced enough to address this. I don't think we can assume that EWB and LWB necessarily will

show the same ontogenetic trend. Appendix Figure 4 clearly shows a more complex mean age related curve for DB than LWB, but the latter will be impacted by the HW-SW colour changes, so it is difficult to judge this specifically. To partly assess this, an age dependent spline is now also used in the analysis presented in Figure 6 and 7. See above comments for Reviewer 1.

L146-147 – This is a fairly short validation period. Why not choose an equally long calibration and validation period which has been a common approach in other similar studies? How sensitive are the results to this choice?
The justification for these periods was not made and this should be clarified. As much of the TR data in Alaska experience calibration and validation issues, specifically in the recent period (so called Divergence), this approach was used for the initial calibration/validation tests. There will certainly be some sensitivity to the periods used, but it should not be forgotten that the full period calibration (1901-2010) was validated against the 19th century data (Figure 7), so multiple periods have been used.

L244-245 acknowledges that this may be an issue. For example, would a different period affect the significance of any results in Table 4?
We believe that consistency of the method for testing each of the parameters and their combinations is more important than using different periods per se. This is a valid point however and certainly the results will change somewhat if different periods were used. However, this was not the aim of the paper and the results of Table 4 and 5 must be pooled together to derive an objective assessment of the varying strengths and weaknesses of these different parameters.

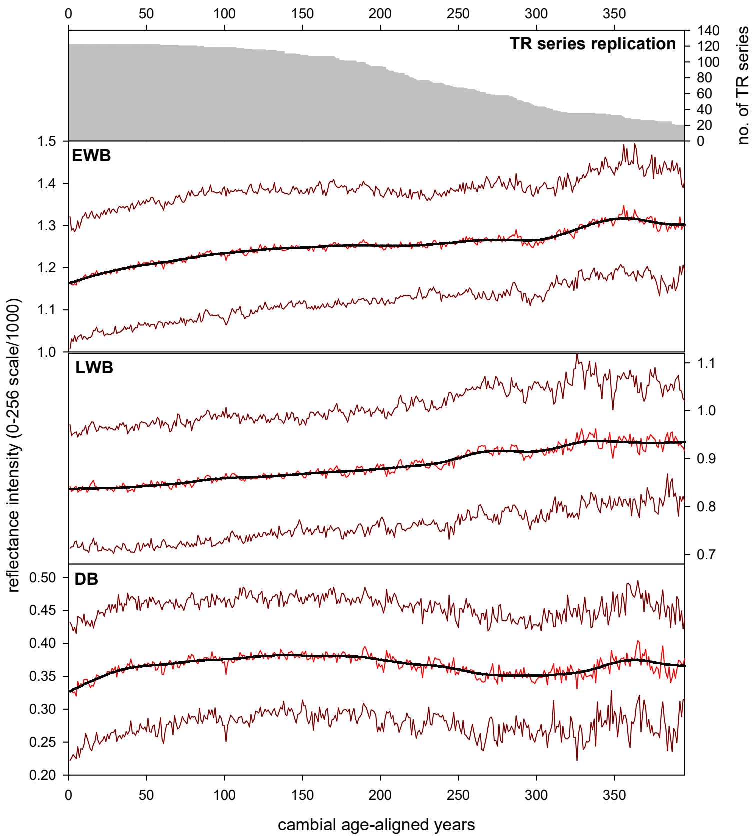
L147-148 and top panel in Figure 8 - Why not perform a nested PC reconstruction? By excluding even just one or two of the shortest sites this reconstruction could be extended to the mid or early 18th century, providing more information for comparison with the other reconstructions.
It is not the specific aim at this time to derive a reconstruction specifically. This paper details multiple experiments to explore the utility of these different parameters for climate reconstruction and to highlight potential parameter specific biases. It is not clear why a nesting approach would change the conclusions already discussed in the paper.

164-173 - Was the variance stabilized to account for changing replication / r-bar through time? Even just a look at the changing variance with replication in Figure S4 in the supplementary material suggests that this should be performed. Either I have missed this or it is not stated in the methods section.
Yes – sorry – variance stabilisation was performed and this should be mentioned in the methodology.

L171 – Figure S4 suggests that linear detrending may not be the most appropriate choice to detrend for example the DB data. Based on the initial increase in the juvenile period of growth in the DB results, perhaps a somewhat more flexible detrending alternative could be explored to account for this initial increase? Additionally, Figure S4 raises some

intriguing questions about non climatic trends in the BI data. I suggest that a fourth panel showing results for EWB should also be shown here. It appears that the initial juvenile trend is present in the EWB only. Does this not suggest that the EWB should not be used to 'correct' LWB in this initial ~30-50 year period?! Because it appears that this initial trend is not present in the LWB data, but is introduced into DB by the EWB data making it necessary to then remove this trend again from the DB series. Due to the experimental nature of this TR variable, several methodological considerations such as the one discussed here remain unaddressed and have not been explored in this study. Considering the nature and aims of this manuscript, I would not expect a detailed evaluation of DB, however, methodological issues such as this and the need to explore them further should at least be acknowledged and more clearly highlighted in the text.
Below is an extended Figure S4 which now includes the EWB age aligned curve. The reviewer is correct that there remains much ambiguity w.r.t. trends as a function of age. We agree that linear functions may not represent the best "fit" for detrending and reviewer 1 mentioned this issue as well.

We have now added an age dependent spline option into the mix. See previous figures.



Is it reasonable to develop a RW + DB chronology / reconstruction considering the difference in the seasonal response of RW and BI data and the acknowledgement that RW may not be primarily controlled by summer temperatures in this region (L349-351 and L379-380)?

Please see the results from the 1st differenced correlation response function analysis results. Yes – we believe that these parameters can be combined although please note that the reconstruction presented in Figure 7 are single parameter only.

L187-189 – Maybe already refer to Figure 2 at the end of the first sentence.

This section focusses only on the common signal and not the climate signal expressed by it. Figure 2 is relevant for the next section.

Perhaps specifically state somewhere in section 3.3 that JJAS was selected for further analysis for BI (presumably because this is the optimal season).
Yes – this can be done.

L196-198 – What might be the reason for such a broad seasonal response in the RW data in this region? Has this been discussed in any of the previous studies (e.g. Wilson et al. 2007 or Wiles et al. 2014)?
Yes – this was discussed in previous papers and is not relevant to the current paper. Please NB the 1st differenced CRFA results which shows that this extended season is still seen for the Feb-August season.

L216 – Just an observation, it is a little surprising that RW agrees more strongly with DB and LWB than DB does with LWB.
The primary author has noted this for multiple species/locations that DB often shows similarities to RW, but with lower 1st order autocorrelation. Remember that all of these parameters should be expressions of summer temperatures so they should show some similarity. This issue needs specific attention at some later stage but is beyond the aim of this current paper.

L219-220 - Maybe worth clarifying here that EWB would not be expected to contain an actual climate signal in the first place.
I do not think we know that specifically and work by Reviewer 1 focussing on the NH Schweingruber work clearly shows a relationship between minimum earlywood density and spring temperatures. It is however, not as strong as MXD with late summer temperatures.

L241-242 – Is it possible that the failed validation could be related to the quality of instrumental data for this early period?
That is indeed likely and we can expand on the discussion in this regard.

L278-280 - This may simply indicate that the use of LWB and EWB to calculate DB is an imperfect procedure. It would be necessary to look at data from other locations (and also other species) to identify whether the DB trends in Figure S4 actually reflect inherent properties related to age (and should therefore be detrended) or are related to other factors. This is an important issue and it is unfortunate that this paper does not or cannot investigate this type of issue in more detail. But perhaps at least a bit more discussion could be included in relation to this?
We like to think that this paper has explored this issue. There is a clear bias in the LWB data which appears partly (completely?) minimised in the use of the DB variant. Yes – clearly this needs more testing, but at the same time, the results may become less ambiguous as more data are added.

L294-295 - The results in Figure S4 already indicate that a linear detrending curve can lead to a serious underestimation in the early parts of the series so the poor performance of LINsf is not surprising.

Yes – but the LINres is fine, so I think we are seeing an issue of the SF approach rather than mis-fitting of a detrending function per se.

L314-317 - But as discussed in the text (and in relation to the results in lines 286-291) it is apparent that LWB is inherently biased. So why even consider it as a feasible option?
Because this is a paper exploring the strengths and limitation of these parameters. Surely we need to show that these data are biased in the low frequency domain. The high frequency signal is however arguably better than DB and don't forget that LWB does in fact pass most validation tests so this is a tricky issue to assess.

L330 – Maybe include '(Figure 6 and 7)' in the bracket since this reconstruction appears in both figures and is discussed in relation to Figure 7 in the previous paragraph.
Yes – can easily add this in.

L329-331 – It may be worth stating here something along the lines that the 'best performing' PC and extended reconstructions are shown here and compared with Wiles and the glacial advance record - i.e. state the reason why these reconstructions are shown in Figure 8.
Yes – can easily add this in.

L358 – Or for species which do not have this colour difference to begin with.
Yes – can easily add this in.

Figure 3 – It is interesting that LWB calibrates and validates more strongly than DB in terms of the strength of the relationship. Why might this be the case? Could this difference be related to replication?
No – replication is the same – this appears to be related to the fact that DB portrays summers temperatures more weakly than LWB – EWB has a weak climate signal and may impact DB when it is calculated. This can be clarified further in the paper and does lead to the possibility that the band-pass approach to calibration used by Rydval et al. (2016/17) could be a realistic approach to dendroclimatic reconstruction in this region. However, this was not the focus of this paper which wanted to highlight the strength and weaknesses of the different parameters. We can add more discussion on this issue.

Figure 6 - It is somewhat difficult to identify the trend of the LINres curve for LWB and especially DB - is it possible to improve the visibility of these curves? Also, in the figure caption (L596-597) maybe consider changing 'low plots' to 'lower set of plots'.
As many of the time-series are very similar, it is really not possible to address the first point in any meaningful well. However, we are happy to edit the caption accordingly.

Figure 7 – Why not also show r2 rather than r for the 1850-1900 period?
We can do this if required but it does not change anything w.r.t. interpretation.

Table 1 - Is there any real meaning in including N-EPS information for EWB data? Presumably these data do not (or should not) contain any common climatic signal and

there would be no point in developing a chronology from these data that would be of much use.

Why would EWB not portray a common signal? We disagree with this comment.

Table 2 – Minor detail, but why not arrange the site order from west to east? Table 3 – Why is EWB positively (though weakly) correlated with DB?

Completely agree. The order can be easily changes to a more logical east-west order as already presented in Table 3.

Technical corrections:

L32 – affecting instead of effecting

Agreed – this can be corrected when "allowed" to edit the paper for final publication

L71 – 'as they are a measure of ...' instead of 'as they measure' may be a more accurate statement.

Agreed – this can be tweaked when "allowed" to edit the paper for final publication

L109 – reconstructions instead of reconstruction

Agreed – this can be corrected when "allowed" to edit the paper for final publication

L126 – Please specify the exact calibration target type as there are different versions. IT8.7/1 is a transparency target whereas IT8.7/2 is a reflective target. I assume that the latter was used.

Yes - IT8.7/2 – text can easily be edited.

L211-212 – Consider rewording 'potentially optimal' to something along the lines of 'more optimal compared to a PC approach'.

We prefer the current wording as a PC approach would not be possible with a sub-fossil extension from one location.

L218 – Consider specifically stating that the correlation between EWB and LWB is not significant.

In actual fact the correlation is significant – but weak. Current wording is preferred.

Also, 'of' missing in 'the utilization DB to'.

Reviewer 1 already flagged this sentence as being clunky. We can revise.

L288 – change 'particular' to 'particularly'

Agreed – this can be corrected when "allowed" to edit the paper for final publication

L637 – Is there a better word than 'dominant' which could be used here?

Can be changed to "Calibration experiments for the four strongest seasons…."

Blue Intensity based experiments for reconstructing North Pacific temperatures along the Gulf of Alaska

Rob Wilson[1,2]; Rosanne D'Arrigo, [2]; Laia Andreu-Hayles[2]; R. Oelkers,[2]; Greg Wiles[2,3]; Kevin Anchukaitis[4,2] and Nicole Davi[2,5]

1. University of Saint Andrews, Saint Andrews, UK; **2.** Tree-Ring Laboratory, Lamont-Doherty Earth Observatory, Palisades, NY, USA; **3**. The College of Wooster, Wooster, Ohio; **4.** School of Geography and Development & Laboratory of Tree Ring Research, University of Arizona, Tucson, AZ USA **5.** William Paterson University, New Jersey, USA.

To be Submitted to Climate of the Past

Abstract: Ring-width (RW) records from the GOA have yielded a valuable long-term perspective for North Pacific changes on decadal to longer time scales in prior studies, but contain a broad winter to late summer seasonal climate response. Similar to the highly climate-sensitive maximum latewood density (MXD) proxy, the Blue Intensity (BI) parameter has recently been shown to correlate well with year-to-year warm-season temperatures for a number of sites at northern latitudes. Since BI records are much less labor intensive and expensive to generate than MXD, such data hold great potential value for future tree-ring studies in the GOA and other regions in mid-to-high latitudes. Here we explore the potential for improving tree-ring based reconstructions using combinations of RW and BI-related parameters (latewood BI and delta BI) from an experimental sub-set of samples at eight mountain hemlock (*Tsuga mertensiana*) sites along the GOA. This is the first study for the hemlock genus using BI data. We find that using either inverted latewood BI (LWB_{inv}) or delta BI (DB) can improve the amount of explained temperature variance by > 10% compared to RW alone, although the optimal target season shrinks to June-September, which may have implications for studying ocean-atmosphere variability in the region. One challenge in building these BI records is that resin extraction did not remove colour differences between the heartwood and sapwood, so long term trend biases, expressed as relatively warm temperatures in the 18th century, were noted when using the LWB_{inv} data. Using DB appeared to overcome these trend biases resulting in a reconstruction expressing 18th-19th century temperatures ca. 0.5°C cooler than the 20th/21st centuries. This cool period agrees well with previous dendroclimatic studies and the glacial advance record in the region. Continuing BI measurement in the GOA region must focus on sampling and measuring more trees per site (> 20) and compiling more sites to overcome site-specific factors affecting climate response and using sub-fossil material to extend the record. Although LWB_{inv} captures the inter-annual climate signal more strongly than DB, DB appears to better capture long term secular trends that agree with other proxy archives in the region. Great care is needed, however, when implementing different detrending options and more experimentation is necessary to assess the utility of DB for different conifer species around the Northern Hemisphere.

Deleted: [1]
Deleted: [1]
Deleted: [1]
Deleted: [1]
Deleted: [1]
Deleted: [1]
Deleted: Tree-Ring Laboratory, Lamont-Doherty Earth Observatory, Palisades, NY, USA;
Deleted: University of Saint Andrews, Saint Andrews, UK;
Formatted: Space After: 0 pt
Deleted: ***PAGES Special Issue:***
Deleted: Climate in the Gulf of Alaska (GOA) reflects large-scale ocean-atmosphere variability of the North Pacific climate system.
Deleted: express
Deleted: expensive and
Deleted: at
Deleted: highlight
Deleted: (LWB)
Deleted: (DB)
Deleted: from
Deleted: such
Formatted: Subscript
Deleted: is not as broad and changes
Deleted: However, o
Formatted: Subscript
Deleted:
Deleted: e
Formatted: Subscript
Deleted: but g
Deleted: needed

Keywords: Blue Intensity, Gulf of Alaska, Tree Rings, Reconstruction, North Pacific; **Short Title**: Gulf of Alaska Blue Intensity Tree-Ring Temperature Reconstruction

1. Introduction

The climate of the Gulf of Alaska (GOA) is strongly influenced by the atmosphere-ocean variability of the North Pacific sector (e.g. the Pacific Decadal Oscillation, Mantua et al. 1997), with profound socioeconomic implications for the region (Ebbesmeyer et al. 1991). The variability of such synoptic climate phenomena is more strongly expressed in winter. Ring-width (RW) data measured from montane treeline conifer trees in the GOA region often express a broad seasonal response window (e.g. January-September, Wilson et al. 2007; February-August, Wiles et al. 2014), which has allowed such data to provide information on cold season synoptic dynamics covering almost two thousand years (Barclay et al. 1999, D'Arrigo et al. 2001, Wiles et al. 2004 and 2014, Wilson et al. 2007).

Maximum-latewood density (MXD) measurements have yielded long records of past summer temperatures for many regions in the northern mid-to-high latitudes (e.g. Schweingruber 1988, Briffa et al. 2002, Anchukaitis et al. 2013, Schneider et al. 2015), but such records do not yet exist for the GOA. MXD series are particularly desirable as such records often have stronger correlations with temperatures than RW and result in climate reconstructions with better skill and spectral fidelity (Anchukaitis et al. 2013, Esper et al. 2015, Wilson et al. 2016, Anchukaitis et al. 2017). This is partly because RW chronologies typically exhibit higher autocorrelation and lagged memory effects than MXD (Briffa et al. 2002; Anchukaitis et al. 2012), but also because RW may potentially integrate other ecological signals (e.g. disturbance and stand dynamics) which can

Deleted: However, t

Deleted: for

Deleted: express

Deleted: coherence

obscure the climate signal (Rydval et al. 2015). Yet, only two millennial-length MXD records are currently published for all of north-western North America (Icefields, British Columbia (BC), Canada - Luckman and Wilson 2005; Firth River, Alaska - Andreu-Hayles et al. 2011, Anchukaitis et al. 2013) and no MXD data have been generated to date for the entire GOA. This situation partly relates to the expensive and labor intensive nature of MXD measurement, but also because the wood of mountain hemlock (*Tsuga mertensiana*), a dominant conifer species in the GOA, is rather brittle and does not lend itself well to standard sample preparation for MXD measurement.

Deleted: northwestern

Deleted: becuase

Deleted: the

To help meet the need for additional climatically-sensitive density records from north-western North America, we present herein an exploration of novel Blue Intensity (BI) parameters measured from scanned images of tree core samples from the GOA. Minimum latewood blue intensity (LWB) has recently been shown to have strong similarities to MXD, and is much cheaper and simpler to generate (McCarroll et al. 2002; Björklund et al. 2014, 2015; Rydval et al, 2014; Wilson et al. 2014, 2017). LWB is closely related to MXD as they both measure similar wood properties (combined hemicellulose, cellulose and lignin content related to cell wall thickness), and both are well correlated with warm-season temperatures (Campbell et al. 2007; Björklund et al. 2014, Rydval et al. 2014, Wilson et al. 2014). This correspondence between BI and temperature has recently been shown to hold true for several locations and tree species, including Scots pine (*Pinus sylvestris*) in Scotland, UK (Rydval et al. 2014) and Sweden (Björklund et al. 2014, 2015), Caucasian fir (*Abies nordmanniana*) in the Northern Caucasus' (Dolgova 2016), Stone pine (*Pinus cembra*) in Austria (Wilson et al. 2017), Engelmann spruce (*Picea engelmannii*) from the Canadian Rockies, British Columbia, Canada (Wilson et al. 2014) and our own analyses of white spruce (*Picea glauca*) in north-western North America (Andreu-Hayles et al., ms. in prep.). Although BI

Deleted: northwestern

Deleted: easier

Deleted: northwestern

often requires larger sample sizes than MXD to improve signal strength (Wilson et al. 2014), this is not a concern due to the low cost of the method.

The greatest limitation of LWB, however, is that any colour variation that does not represent year-to-year climate-driven cell wall thickness changes will bias the resultant raw reflectance measurements. For example, some conifer species (including Scots pine and mountain hemlock) show a clear sharp or transitional colour change from the heartwood to sapwood, which, even after resin extraction using ethanol or acetone, can still impose a systematic change in reflectance around the heartwood/sapwood transition (Rydval et al. 2014; Björklund et al. 2014, 2015). Further colour variations, often seen in dead but preserved snag or sub-fossil wood, can also result in systematic biases when combined with data measured from living samples (Björklund et al. 2014, 2015; Rydval et al. 2014). Björklund et al. (2014) proposed a potential solution to the heartwood/sapwood colour bias issue by effectively detrending the LWB measurements by removing the inherent common colour changes of the earlywood and latewood (i.e. those related to heartwood/sapwood colour change). This is accomplished by subtracting the raw LWB value from the maximum blue reflectance value of the earlywood (EWB) for each year. The resulting new parameter, delta blue intensity (hereafter referred to as DB), should theoretically be less biased by such non-climatic related colour changes. Although Björklund et al. (2014, 2015) presented compelling results using Scots pine in Sweden, DB has not yet been tested elsewhere or on any other species.

Deleted: was

Deleted: done

Deleted: undertaken

Deleted: referred to as

Deleted: DB

Deleted: the method

Deleted: Importantly, DB can only theoretically work if the inter-annual signal between EWB and LWB is weakly correlated. If the correlation between these two parameters is high, then the method of deriving DB may remove the specific climate signal of interest.

Finally, although BI based variables hold great promise as an alternative proxy to MXD, another potential concern is the possibility that reflectance based measurements may not capture low

Deleted: their potential inability to frequency information related to long term-climate changes. Wilson et al. (2014), working with Engelmann spruce from British Columbia, which does not have a visual colour difference between the heartwood and sapwood, urged caution as both the MXD and LWB parameters were sensitive to different detrending options and there was some indication that LWB could not capture as much low frequency information as MXD. However, this observation could not be fully addressed due to the relatively short instrumental record in British Columbia.

Deleted: express

In this paper, building upon previous RW based research (Wilson et al. 2007, Wiles et al. 2014), we measure BI variables (EWB, LWB and DB) from multiple sites in the GOA to evaluate: (a) whether BI can improve on previous RW-only based reconstructions, and (b) whether meaningful low frequency information can be gleaned from these data by exploiting the long monthly instrumental record from Sitka, Alaska back into the mid-19th century to validate secular trends in the TR data.

2. Methods and Analysis

For this exploratory study, BI measurements were made on a subset (ca. 15 single tree cores per site) of crossdated core samples collected over the past few decades from living mountain hemlock (*Tsuga mertensiana* Bong. Carrière) trees located at eight sites near altitudinal treeline (approximately ~300-400 meters above sea level) along the GOA (Table 1, Figure 1). Data from these and additional sites were used previously to create coastal GOA RW based temperature-related reconstructions (D'Arrigo et al. 2001, Wilson et al. 2007, Wiles et al. 2014).

Deleted: -

Deleted: around

The tree core samples were immersed in acetone for 72 hours to remove excess resins in the wood (Rydval et al. 2014) and then finely sanded to 1200 grit to remove marks and abrasions prior to scanning. An Epson V850 pro scanner, using an IT8.7/2 calibration card in conjunction with Silverfast scanning software, was used to scan the samples at 2400 dpi resolution. Raw EWB and LWB variables were measured using CooRecorder 8.1 software (Cybis 2016 - http://www.cybis.se/forfun/dendro/index.htm), which has capabilities to acquire accurate reflectance intensity RGB colour measurements from scanned wood samples (see Rydval et al. 2014). DB values were calculated within CooRecorder by subtracting the raw LWB values from the raw EWB values for each year. Since raw LWB is negatively correlated to MXD (high density 'dark' latewood = low reflectance), values were inverted following the method detailed in Rydval et al. (2014) to allow for LWB (hereafter denoted as LWB_{inv}) to be detrended in a similar way to MXD (see also Wilson et al. 2014). The nature of the DB calculation results in this parameter being positively correlated with LWB_{inv}, so these data could also be theoretically detrended in a similar way. It should be noted that Björklund et al. (2014) proposed that LWB_{inv} should be referred to as maximum latewood blue absorption intensity.

As the mean sample length (Table 1) for all sites was > 200 years, for initial experiments comparing the different tree-ring (TR) variables, the RW, LWB_{inv}, EWB and DB data were detrended using fixed 200-year cubic smoothing splines (Cook and Peter 1981) to retain the interannual to multi-decadal signal and minimize any potential lower frequency biases due to heartwood/sapwood colour changes. The variance of the site and regional composite chronologies were temporally stabilized using techniques detailed in Frank et al. (2007a). These chronology versions were assessed by (1) signal strength statistics: both common signal (via mean inter-series

Deleted: IT-8
Deleted: state-of-the-art
Deleted: LWB
Formatted: Subscript
Deleted: inverted
Formatted: Subscript
Deleted: F
Formatted: Subscript
Deleted:

correlation – RBAR) and expressed population signal (EPS - Wigley et al. 1984) statistics, (2) between variable correlation, (3) between site coherence using a rotated principal component analysis (PCA, varimax rotation using correlation matrices with eigenvectors retained with an eigenvalue > 1.0) and (4) climate response derived by correlations between regional composite TR variable mean series and the dominant PC scores against monthly and season variables of temperature (CRU TS 3.24 (Harris et al. 2012): 57-61°N / 153-134°W).

The 200-year spline chronology versions were also used to explore calibration (1901-1960) and validation (1961-1989) based principal component regression reconstruction experiments using the CRU TS data. The 1961-1989 period was specifically used for validation as many tree-ring width based temperature sensitive chronologies in Alaska do not track recent temperature trends well – a phenomenon often referred to as the "Divergence Problem" (D'Arrigo et al. 2008). For the PCA, a reasonably replicated common period (1792-1989) was used where tree series replication was > 5 trees. All site chronologies are replicated with > 10 trees from 1792 except for JM and SR (see Table 1) where replication is 6 and 5, respectively. Reconstruction validation was performed using the Pearson's correlation coefficient (r), the Reduction of Error (RE) and the Coefficient of Efficiency (CE - Cook et al., 1994). Further validation was performed over the 1850-1900 period using the gridded BEST instrumental data (Rohde et al., 2012), extracted for the same region as the CRU TS (57-61°N / 153-134°W), after these data were scaled to the CRU TS data over the 1901-2015 period. CRU TS and BEST are compared (Supplementary Figure S1) to the original GOA 5-station mean records used in Wilson et al. (2007) to confirm that the gridded products are good representations of the regional temperature signal. The higher variance of the pre-1950 period in the 5-station mean is related to the fact that variance stabilization (Frank et al.

2007a) was not performed when this mean series was originally developed (Wilson et al. 2007) and is therefore likely a less robust measure of GOA temperatures than the gridded products.

Finally, to improve overall expressed signal strength and explore the potential of reconstructing robust low frequency temperature changes in the region, the data from each of the eight sites were pooled to derive GOA regional composite records for each of the TR variables. These pooled composite variable datasets, with their greater overall replication, allowed detrending experiments to be performed to ascertain the sensitivity of the final parameter chronologies to different detrending choices. Specifically, RW detrending experiments were performed using (1) STD: negative exponential function or negative or zero slope linear function detrending via division; (2) NEPT: negative exponential function or negative or zero slope linear function detrending via subtraction after power transformation of the raw RW data (Cook and Peters 1997); and (3) RCS: single group regional curve standardization (RCS - Briffa et al., 1996; Esper et al., 2003; Briffa and Melvin 2008) detrending via division. The regional age-aligned curve was smoothed using a cubic smoothing spline (Cook and Peters 1981) of 10% the series length. For each of these three approaches, the 'Signal-Free' (SF - Melvin and Briffa 2008) approach to detrending was also used. Finally, the composite chronologies, also using the SF approach, were also derived using the age dependent spline (ADS) approach introduced by Melvin et al. (2007) to track more complex growth trends that may not be captured well with the STD, NEPT and RCS approaches. These different detrending options resulted in an ensemble of 7 different RW composite chronologies. For LWB_{inv} and DB, as they theoretically should behave more like MXD, which often has a decreasing linear trend, detrending was performed using (1) LINres: negative or zero slope linear function detrending via subtraction - with and without the SF approach; (2) RCSres: single group

Deleted: utilized
Deleted: 6
Deleted: expresses
Formatted: Subscript
Deleted: only two methods;

RCS detrending via subtraction - with and without the SF approach; and (3) ADSsf: the signal free age dependent spline approach. Overall, for LWB_{inv} and DB, five chronology variants were developed for analysis.

3. Results and Discussion

Common signal within the network

RW has the strongest common signal with a median overall RBAR of 0.44 (8 site range: 0.33 – 0.49 - Table 1), whereas LWB_{inv} and DB both have weaker RBAR values of 0.24. EWB shows the weakest common signal with a median RBAR from the 8 sites of only 0.12. In order of decreasing between-series common signal, the number of trees needed to attain an EPS of 0.85 are 7 (RW), 18 (LWB_{inv} and DB), and 41 (EWB) for each variable respectively. On average, therefore, except for RW, actual replication for the reflectance based parameter chronologies are often lower than would be ideally needed to attain a robust expressed population signal. This is important to keep in mind as it is likely that the experimental calibration results presented herein will improve as replication is increased.

The weak signal strength in EWB compared to RW, LWB_{inv} and DB is also reflected in the PCA. The leading PC for RW, LWB_{inv} and DB explains 59%, 53% and 57% of the overall variance, respectively, while just 39% is explained by the EWB PC1. In general, the loadings (based on a varimax rotation) of the chronologies on each PC for each variable are related to the geographical locations across the GOA with PC1 representing the eastern sites and PC2 the western ones (Figures 1 and 2). A similar spatial distribution of loadings was noted in Wilson et al. (2007) using RW data from 31 living sites across the GOA.

Deleted: . As with the RW data, the SF approach was also performed leading to f

Deleted: our

Deleted: for both LWB and DB

Formatted: Subscript

Formatted: Subscript

Deleted: series

Deleted: LWB

Deleted: TR

Deleted: Rydval et al. (2014) showed that as the within tree common signal was much weaker for LWB than RW, the between tree common signal improved more for LWB than RW as multiple radii from the same tree were measured (i.e. up to 3). For this exploratory analysis, only a single series was measured per tree, and therefore we hypothesise that the EPS of BI based chronologies would improve markedly, compared to RW, if at least 2 radii were measured from each tree.

Deleted: LWB

Deleted: LWB

Seasonal temperature sensitivity

EWB contains a weak response to summer temperature variability with almost no late summer temperature signal (Figure 2) although some significant correlations (r = ~0.3 - 0.4) are found with May and previous October/November temperatures (supplementary Figure S2). Correlations with seasonal temperatures, after 1^{st} differencing, identifies no significant response (supplementary Figure S3). In agreement with previous work (Wilson et al. 2007; Wiles et al. 2014), RW correlates well with a broad range of summer seasons (Figure 2), showing positive correlations for nearly all months from January through to September (Supplementary Figure 2) with June returning the strongest correlation. Correlations do weaken when the data are 1^{st} differenced (supplementary Figure S3), but the Wiles et al. (2014) RW composite still retains a strong response with February-August temperatures although for the other RW based time-series, the summer season shows the strongest coherence. LWB_{inv} and DB, show a weaker response with the late winter/spring months compared to RW and strongest correlations with June, July and August (Figure 2). These observations were expected as LWB_{inv} and DB should express similar growth/climate response properties to MXD.

For the RW, LWB_{inv} and DB data, there appears to be a geographical difference in response with PC1 (eastern sites) showing stronger seasonal (Figure 2) and monthly (supplementary Figure S2) correlations with temperature than PC2 (western sites). However, correlations of the individual site chronologies for each TR variable (Table 2) against June-September temperatures (optimal season for reconstruction – see later) suggest that there is a degree of variability of the individual sites' response to summer temperatures across the GOA. As PC2 is weighted more towards the

Formatted: Superscript
Deleted: a little
Formatted: Superscript
Deleted:
Formatted: Subscript
Deleted: on the other hand,
Deleted: s
Deleted: A
Deleted: LWB
Deleted: , these observations are not surprising
Formatted: Subscript
Deleted: T
Deleted: . This spatial pattern of response is also expressed in the RW data (Figure 2).

TBB site (see PCA loadings in Figure 2 for RW, LWB_{inv} and DB) which correlates weakly with JJAS, it is therefore not surprising that this PC correlates weakly with summer temperatures. After 1st differencing, however, these regional differences disappear (supplementary Figure S3) suggesting there are potential post-detrending trend biases in the chronologies weighted on PC2.

It is important to note that the correlation of the mean composite chronologies with summer temperatures (Figure 2 and especially supplementary Figure S3 after a 1st differenced transformation) are stronger than the PC1 results. This suggests that a regional mean composite approach is potentially optimal in the context of deriving a GOA wide reconstruction which can be extended in the future by data generated from sub-fossil samples.

The positive correlation of RW, LWB_{inv} and DB to summer temperatures (Figure 2 and Table 2) is also reflected in the inter-correlation between these different variables (Table 3). RW agrees most strongly with DB, followed by LWB_{inv}. EWB, unsurprisingly, has the weakest relationship with the other 3 TR variables. Hereafter, due to the poor signal strength and weak climate signal, the EWB data were not used for further analysis, except in the DB calculations.

Calibration/validation experiments

Calibration and validation statistics for various PC regression variable combinations for several summer target seasons are detailed in Table 4 along with results using the GOA RW composite of Wiles et al. (2014). Firstly, calibration of Wiles et al. (2014) to the CRU TS 3.24 data (February – August) over the 1901-1989 period (r^2 = 0.33) is stronger than the new RW GOA composite (r^2 = 0.27) which also shows a significant trend in the model residuals. This residual trend possibly

Deleted:

Deleted: However,

Deleted: results

Formatted: Superscript

Deleted: marginally

Deleted:

Formatted: Subscript

Formatted: Subscript

Deleted: Importantly, the correlation between EWB and LWB is weak which is the theoretical ideal for the utilization DB to minimize potential heartwood/sapwood colour change biases (Björklund et al. (2014).

Deleted: alone

Deleted: was

Deleted:

reflects the fact that there could be a longer term trend missing in the RW data due to the use of 200-year spline detrended chronologies when compared to the RCS processed version of Wiles et al. (2014). Also, the slightly weaker results for the new RW data likely reflect lower replication in the current study compared to Wiles et al. (2014).

The strongest calibration $_{a}r^{2}$ values for each BI parameter over the 1901-1960 period are 0.49 and 0.47 for LWB$_{inv}$ and DB respectively for the JJA season although DB fails validation with negative RE and CE values over the 1961-1989 period. Minimal model improvement is gained by including RW data. RW+LWB$_{inv}$ calibrates best ($_{a}r^{2}$ = 0.49) with JJA while RW+DB explains more temperature variance for MJJAS ($_{a}r^{2}$ = 0.51). However, in both cases, validation RE and CE are negative. Focussing on the full period (1901-1989) calibration, strongest results are found for the JJAS season for all parameter options with $_{a}r^{2}$ values of 0.27 (RW), 0.43 (LWB$_{inv}$), 0.38 (DB), 0.38 (RW+LWB$_{inv}$) and 0.39 (RW+DB) with no 1st order autocorrelation observed for any version. Importantly, only the RW+DB version shows no significant linear trend in the model residuals. The full period (1901-1989) calibrated reconstructions (Table 4) for each of the variable options are presented in Figure 3 along with independent validation (1850-1900) with the BEST gridded data. All parameter iterations fail validation (negative CE values) except for RW+DB which returns positive RE (0.57) and CE (0.19) values. Overall, using this subset of samples from these 8 sites, the calibration results (Table 4 and Figure 3) indicate that BI based parameters explain more temperature variance than using RW alone. However, the fidelity of the resultant reconstructions appears sensitive to the periods of calibration and validation used and it is not clear which of these parameters best represent longer term secular change as the chronologies were limited in the frequency domain by using a fixed 200-year spline detrending option.

Deleted: low frequency

Deleted: of

Deleted: generally

Formatted: Subscript

Deleted: s

Deleted: noted

Deleted: found

Deleted: Only

Deleted: , however,

Deleted: assessing

Deleted: and as the chronologies were limited in the frequency domain by using a fixed 200-year spline detrending option, it is not clear which of these parameters best represent longer term secular change

The large-scale climate signal expressed by these data is illustrated by comparing the RW+DB JJAS reconstruction with gridded land/sea HadCRUT4 (Morice et al. 2012; Cowtan and Way 2014 – Figure 4a) and land only CRU TS 3.24 (Harris et al. 2012 – Figure 4b) temperatures for the GOA and North Pacific sector. Although the spatial correlations are stronger towards Juneau and Sitka (see Figure 1 for locations) in the east of the region it is clear that these new data represent the temperature variability of the wider GOA region and North Pacific. Continued measurement of BI based parameters from sub-fossil samples taken from across the GOA will allow long term summer temperature variability to be derived for at least the last millennium which will complement the long RW based temperature reconstructions expressing a broader seasonal window (Wilson et al. 2007; Wiles et al. 2014).

Potential low frequency bias

The main potential limitation to the use of BI based TR variables such as LWB is concerned with low frequency trend biases related to wood colour change. Mountain hemlock, in general, shows darker heartwood and lighter sapwood which resin extraction appears to only minimise but not entirely remove. However, this colour change is not a sharp transition and is expressed in raw EWB and LWB measurements as a steady increase in reflectance intensity. Non-detrended mean composite chronologies of EWB and LWB for the whole GOA region (Figure 5) clearly show the impact of the heartwood/sapwood colour change with increasing intensity values through time (see also Supplementary Figure S4 for a single tree example), especially since the late 18th century. In contrast, MXD generally shows a linear decreasing trend with increasing cambial age (Esper et al. 2012). If LWB is indeed a comparable (but inverted) TR variable to MXD as a measure of

Deleted: well

Deleted: each end of

Deleted: 3

Deleted: However

latewood anatomical density properties, then we would expect, therefore, an increasing trend in raw LWB values. Figure 5 therefore poses a potential "mixed-signal" conundrum as the observed trend in the GOA raw mean LWB composite will incorporate the secular climate signal, the true age-related trend of changing latewood density, and the heartwood/sapwood colour change bias. Although using DB can theoretically overcome the colour bias issue, it has not been explored in any detail beyond the original concept papers (Björklund et al. 2014, 2015). The mean DB non-detrended GOA chronology (Figure 5) has minimal long term trends, which could suggest that the colour change bias has been removed or at least minimised.

Mean cambial age-aligned curves of the EWB, LWB and DB data show very distinct trends (Supplementary Figure S5). LWB appears to show a general linear increase in values – a trend that would be expected if LWB indeed does reflect similar wood properties (inversely) to MXD. DB, however, has a more complex mean growth curve, essentially reflecting trends in the EWB data, and shows an initial increasing juvenile trend for ~50 years, a period of stabilisation and then a decreasing trend from about ~200 to 300 years. These different age-aligned curves highlight that different detrending options may well be needed for these different TR variables.

A range of credible options for detrending the RW, LWB_{inv} and DB GOA regional composite data are presented in Figure 6. The outcome for the RW data appears extremely consistent even when using STD vs RCS based methods. However, the LWB_{inv} and DB chronologies are much more sensitive to the detrending method used. Compared to RW and DB, all LWB_{inv} chronology variants show above zero index values in the 18^{th} century, which likely reflects the low reflectance bias of the darker heartwood compared to the sapwood because the LWB_{inv} data have been inverted. The

Deleted: both
Deleted: scale
Deleted: these
Deleted: latter
Deleted: s
Deleted: expresses
Deleted: 4
Deleted: methodological
Deleted: choice of
Formatted: Subscript
Deleted: LWB
Deleted: extremely
Deleted: LWB
Deleted: 7
Deleted: LWB

RCS versions appear particularly inflated and as LWB$_{inv}$ is positively correlated with summer temperatures (Figure 2), this would result in markedly warm temperature estimates during the LIA compared to the 20th century which is at odds with previous GOA dendroclimatic analyses (Wiles et al. 2014) and the geomorphological record, which indicate substantial cool conditions and glacial advance from the 17th to 19th centuries (Wiles et al., 2004; Solomina et al. 2016). RCS can impart significant low frequency bias when the assumptions and requirements of the method are not met (Melvin and Briffa 2014; Anchukaitis et al. 2013) and as the GOA composite utilises only living trees, this is a far from optimal sample design for this detrending method. For DB, the LINsf version deviates markedly from LINres, RCSres, RCSsf and ADSsf variants with very low values (< -6 standard deviation from 1901-1989 mean) before 1700 followed by a strong linear increase until present. A similar observation was noted in Wilson et al. (2014) where signal free detrending of LWB$_{inv}$ and MXD resulted in much cooler LIA conditions than other detrending approaches.

JJAS GOA summer temperatures back to 1600

The long GOA instrumental record allows for additional assessment of how different reflectance based chronology variants track temperatures back through time. Using the extended BI based regional composite records, further reconstruction experiments against the JJAS season were performed using LWB$_{inv}$ and DB separately (Table 5) by calibrating against JJAS CRU TS3.24 (1901-2010) and separately validating using the BEST data (1850-1900). For the LWB$_{inv}$ data, due to their strong post 1970s decreasing trends (Figure 6), RCSres and RCSsf calibrated poorly (Table 5: r^2 = 0.07 and 0.05 respectively) and validated with negative CE values over the 1850-1900 period. LINres and LINsf explained 41% of the temperature variance while the ADSsf variant explained 47%. All three versions validated reasonably well with positive RE and CE values.

Deleted: LWB
Deleted: s
Deleted: RCSsf
Deleted: LWB
Deleted: especially
Deleted: LWB
Deleted: LWB
Deleted: with
Deleted: The
Deleted: version, however,
Deleted: and

Significant 1st-order autocorrelation (DW range 1.28 to 1.37) and linear trends (LINr range 0.36 to 0.48) were however noted for all model residuals except for ADSsf. For the DB data, calibration was strongest for RCSsf, RCSres and ADSsf with positive RE and CE values for all versions except ADSsf which failed the CE test. The residuals from the RCSsf, RCSres and ADSsf calibrations show no 1st-order autocorrelation although a significant linear trend is observed.

Considering the calibration and validation experiments presented in Tables 4 and 5, our results do not definitively identify whether LWB_{inv} or DB is the optimal BI based parameter for reconstructing past summer temperatures for the GOA region. Part of this ambiguity potentially stems from unknown uncertainties in the 19th century instrumental data, but the sensitivity of the TR parameters to different detrending options (Figure 6) exacerbates the situation. Calibration suggests that the inter-annual based signal of LWB_{inv} is marginally stronger than DB but validation against the 19th century data cannot distinguish between the different parameters. The clear differences between the chronology versions (Figure 6), especially before 1850, have huge implications for understanding past temperatures in the region.

To try and derive a parameter specific view of long term temperature changes for the region, a weighted mean using the five different variants were combined to create a regional average. The r2 values, derived from the 1901-2010 calibration (Table 5), was used as a weighting term to calculate the parameter specific weighted averages which were then calibrated (1901-2010) and validated (1850-1900) in the same way as detailed in Table 5. The resultant weighted LWB_{inv} and DB regional reconstructions report quite different histories of past GOA temperatures (Figure 7). Specifically, the LWB_{inv} reconstruction has temperature estimates from the late 17th to mid-19th

Deleted: was

Deleted: The

Deleted: chronology variants

Deleted: on the whole performed better than their LWB counterparts, with

Deleted: (

Deleted:)

Deleted: calibrating most strongly (r^2 = 0.43 (0.40)) and validating well (RE = 0.48 (0.50), CE = 0.15 (0.18))

Deleted: for

Deleted: both versions

Deleted: however

Deleted: there are no clear

Deleted: that

Formatted: Superscript

Deleted: also

Formatted: Superscript

Deleted: ¶
The best reconstructions using LWB (LINres) and DB (RCSsf), identified using the calibration and verification results (Table 5), represent

Deleted: LWB (LINres)

Deleted: expresses

century warmer than the 1961-1990 mean, while the DB reconstructions exhibits generally cooler conditions. Both reconstructions explain a similar amount of summer temperature variance (LWB$_{inv}$: 43% vs. DB: 40%) and validate well with positive RE and CE values. The regression residuals from both versions have a significant linear trend, but only for DB was no significant 1st order autocorrelation identified in the residuals. From comparison to the instrumental data alone, therefore, one cannot objectively identify which of the two parameter versions is most robust. Wilson et al. (2014) highlighted the difficulties of relying solely on the instrumental data to validate the long-term trend in any reconstruction. Moreover, there could be unknown inhomogeneity issues in early instrumental data series which are difficult to identify which would influence calibration and validation (see Frank et al. 2007b). Therefore, alternative sources of relevant information are recommended for further validation. As the geomorphological record in the region suggests a prolonged period of glacial advance occurred in the GOA up to the early 20th century (Wiles et al., 2004; Solomina et al. 2016) when a substantial retreat started, we hypothesize that the pre-1900 period must therefore have been cooler. This would suggest that the DB based reconstruction is likely more representative of past GOA temperatures than the LWB$_{inv}$ driven one.

Figure 8 presents the RW + DB principal component reconstruction (Figure 3), the weighted LWB$_{inv}$ and DB extended reconstructions (Figure 6), and the Wiles et al. (2014) RW based reconstruction and compares them to the GOA regional glacial advance record (Wiles et al., 2004; Solomina et al. 2016). The LWB$_{inv}$ reconstruction is clearly at odds with the other records with warmer than average temperatures for many periods over the last 400 years and no specific prolonged cooler periods though the LIA. The other TR reconstructions demonstrate centennial and multi-decadal agreement, although the extended DB reconstruction exhibits a smaller

Deleted: (RCSsf)

Deleted: (Table 5)

Formatted: Subscript

Deleted: express

Deleted: is no

Deleted: and it is not

Formatted: Superscript

Deleted: was

Deleted: and f

Deleted: choose

Deleted: although there are arguably less problems with the model residuals when utilising the DB data

Deleted: impact

Deleted: needed

Formatted: Subscript

Deleted: (RCSsf)

Formatted: Not Superscript/ Subscript

Deleted: expressing

Deleted: that have

Deleted: has

amplitude of temperature change between the LIA period and the 20th century. Overall, temperatures in the GOA region were below the 1961-90 norm throughout most of the LIA with temperatures only rising to substantially higher values in the early 20th century. The coldest decadal periods are centred around the 1700s, 1750s, and 1810s. The glacial advance record shows periods of advance through the LIA, peaking at the end of the 19th century. Despite the use of 200-year spline detrended chronologies, the RW+DB reconstruction has a similar amplitude change to the Wiles et al. (2014) record, which was derived from RCS processed RW data. It should be noted also that this RW based reconstruction was calibrated against Feb-August temperatures which has a greater increasing temperature trend (0.81°C/century vs 0.62°C/century) and higher variance (0.79 vs 0.41) than JJAS (calculated using BEST data from 1850-2015), which will influence the amplitude of the reconstruction (Esper et al. 2005).

Deleted: clearly

Deleted: however

4. Conclusions

We have described a set of experimental temperature reconstructions based on RW, LWB_{inv} and DB data measured from eight tree-ring sites along the Gulf of Alaska. Focusing on these data sets, the results demonstrate that inclusion of BI based variables can significantly improve the calibrated variance explained using RW alone by more than 10%.

Formatted: Subscript

RW, LWB_{inv} and DB are strongly correlated with each other (Table 3) but the inclusion of LWB_{inv} or DB shifts the calibrated signal from a broad (February-August, Wiles et al. 2014) season using RW alone to a late summer (JJAS) season. The influence of late winter and early spring temperatures on RW suggest that this variable may, in fact, still be the more optimal variable for studying important synoptic phenomena such as north Pacific variability, which dominates in the

Deleted: LWB

Deleted: LWB

winter months (Wilson et al. 2007).

The LWB_{inv} data, for mountain hemlock, despite calibrating and validating in a similar way to DB, are clearly affected by heartwood/sapwood colour differences which impart a trend bias in the resultant chronologies and reconstructions (Figures 6, 7 and 8). However, this bias may not necessarily always occur for other species showing a heartwood/sapwood colour change which could be removed through traditional resin extraction methods. For the first time since the original concept papers by Björklund et al. (2014, 2015), we have experimented with the DB variable. The resulting reconstruction agrees well with a previous RW based reconstruction (Wiles et al. 2014) and the glacial advance record (Wiles et al., 2004; Solomina et al. 2016) for the region.

Deleted: LWB

Deleted: 7

Deleted: and t

The analyses presented herein must be viewed as a series of experiments to inform dendroclimatologists of possible methodological strategies that need to be considered for improving TR based reconstructions using BI based variables. Specific to the GOA region, but likely relevant to other regions and species, we therefore detail the following recommendations:

Deleted: future

Deleted: blue reflectance

- Although MXD typically has a higher expressed population signal (EPS) strength and climate responses than RW (Wilson and Luckman 2003), signal strength in LWB and DB in GOA hemlock is weaker than RW, so replication needs to be substantially increased (ideally > 20 trees – Table 1) to allow the development of robust chronologies. Rydval et al. (2014) also showed that a substantial improvement in LWB signal strength could be gained by measuring 2 or even 3 radii per tree. Additional assessments of signal strength should be conducted as new species and sites are analysed using BI methods.
- For conifer species with a clear colour difference between the heartwood and sapwood,

LWB_{inv} may likely always contain biased long-term trends. The DB variable could potentially minimize this effect as shown here, but more experimentation with this parameter is needed before it can be commonly used as a solution to the LWB_{inv} colour bias problem. Rydval et al. (2017), using Scots pine, overcame the heartwood/sapwood colour bias by utilising a band-pass approach to calibration, where LWB_{inv} drove the decadal and high frequency fraction of the Scottish temperature reconstruction, while RW drove the low frequency variability. This approach however assumes that (1) RW is predominantly controlled by summer temperatures and (2) meaningful long-term information can be gleaned from RW data, which may not always be the case (Esper et al. 2012).

- There is substantial sensitivity of the final chronologies to varying methodological detrending approaches. Much more exploration of the impact of different detrending choices is needed and it is likely that ensemble based approaches (Wilson et al. 2014) will ultimately be the only way to derive realistic estimates and appropriate detrending based uncertainties bounds. Locations with long instrumental records may help identify more optimal detrending options but care is needed, as it cannot be assumed that the quality of 19th century data is comparable to late 20th/early 21st century data. Utilizing other proxy observations of past climate (e.g. in this case the glacial record) may help further constrain TR estimates of past climate especially when different chronology variants (that validate well) portray quite different past temperature histories.

Acknowledgments. This is a contribution to the PAGES 2k Network [through the Arctic/North America-2k working groups]. Past Global Changes (PAGES) is supported by the US and Swiss

Formatted: Subscript

Deleted: express

Deleted: LWB

Deleted: LWB

Deleted: (not necessarily the case in the GOA)

Deleted: er

Deleted: secular

Deleted: The results presented herein highlighted

Deleted: error

National Science Foundations. We also gratefully acknowledge the National Science Foundation's Paleoclimatic Perspectives on Climatic Change (P2C2) Grant Nos. AGS 1159430, AGS 1502186, AGS1502150 and PLR 15-04134. The detailed reviews from Milos Rydval and Jesper Björklund were very much appreciated. Lamont-Doherty Earth Observatory Contribution No. 8121.

Deleted: and

Deleted: 0000

5. References

Anchukaitis, K.J., Breitenmoser, P., Briffa, K.R., Buchwal, A., Büntgen, U., Cook, E.R., D'arrigo, R.D., Esper, J., Evans, M.N., Frank, D. and Grudd, H., 2012. Tree rings and volcanic cooling. Nature Geoscience, 5(12), pp.836-837.

Anchukaitis, K.J., D'Arrigo, R.D., Andreu-Hayles, L., Frank, D., Verstege, A., Curtis, A., Buckley, B.M., Jacoby, G.C. and Cook, E.R., 2013. Tree-ring-reconstructed summer temperatures from northwestern North America during the last nine centuries. Journal of Climate, 26(10), pp.3001-3012.

Deleted: ¶

Anchukaitis, K.J., Wilson, R., Briffa, K., Büntgen, U., Cook, E.R., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P.J., Linderholm, H., Myglan, V., Osborn, T., Peng, Z., Rydval, M., Schneider, L., Schurer, A., Wiles, G. and Zorita, E. 2017. Last millennium Northern Hemisphere summer temperatures from tree rings: Part II: spatially resolved reconstructions, Quaternary Science Reviews, 163, 1-22, doi: 10.1016/j.quascirev.2017.02.020

Andreu-Hayles, L., D'Arrigo, R., Anchukaitis, K.J., Beck, P.S., Frank, D. and Goetz, S., 2011. Varying boreal forest response to Arctic environmental change at the Firth River, Alaska. Environmental Research Letters, 6(4), p.045503.

Barclay, D.J., Wiles, G.C. and Calkin, P.E., 1999. A 1119-year tree-ring-width chronology from western Prince William Sound, southern Alaska. The Holocene, 9(1), pp.79-84.

Björklund, J.A., Gunnarson, B.E., Seftigen, K. and Esper, J., 2014. Blue intensity and density from northern Fennoscandian tree rings, exploring the potential to improve summer temperature reconstructions with earlywood information. Climate of the Past, 10(2), p.877.

Björklund, J., Gunnarson, B.E., Seftigen, K., Zhang, P. and Linderholm, H.W., 2015. Using adjusted Blue Intensity data to attain high-quality summer temperature information: A case study from Central Scandinavia. The Holocene, 25(3), pp.547-556.

Briffa, K.R., Jones, P.D., Schweingruber, F.H., Karlén, W. and Shiyatov, S.G., 1996. Tree-ring variables as proxy-climate indicators: problems with low-frequency signals. In Climatic variations and forcing mechanisms of the last 2000 years (pp. 9-41). Springer Berlin Heidelberg.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., Jones, P.D., Shiyatov, S.G. and Vaganov, E.A., 2002. Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals. The Holocene, 12(6), pp.737-757.

Briffa, K.R. and Melvin, T.M., 2011. A closer look at regional curve standardization of tree-ring records: justification of the need, a warning of some pitfalls, and suggested improvements in its application. In Dendroclimatology (pp. 113-145). Springer Netherlands.

Campbell, R., McCarroll, D., Loader, N.J., Grudd, H., Robertson, I. and Jalkanen, R., 2007. Blue intensity in Pinus sylvestris tree-rings: developing a new palaeoclimate proxy. The Holocene, 17(6), p.821.

Cook ER, Peters K. 1981. The smoothing spline: a new approach to standardising forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bulletin 41: 45–54.

Cook, E.R. and Peters, K., 1997. Calculating unbiased tree-ring indices for the study of climatic and environmental change. The Holocene, 7(3), pp.361-370.

Cowtan, K. and Way, R.G., 2014. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. Quarterly Journal of the Royal Meteorological Society, 140(683), pp.1935-1944.

D'Arrigo, R., Villalba, R. and Wiles, G., 2001. Tree-ring estimates of Pacific decadal climate variability. Climate Dynamics, 18(3-4), pp.219-224.

D'Arrigo, R., Wilson, R., Liepert, B. and Cherubini, P., 2008. On the 'divergence problem' in northern forests: a review of the tree-ring evidence and possible causes. Global and Planetary Change, 60(3), pp.289-305.

Dolgova, E., 2016. June–September temperature reconstruction in the Northern Caucasus based on blue intensity data. Dendrochronologia, 39, pp.17-23.

Ebbesmeyer, C.C., D.R. Cayan, D.R. McLain, F.H. Nichols, D.H. Peterson and K.T. Redmond, 1991:1976 step in the Pacific climate: Forty environmental changes between 1968-75 and 1977-1984. In: Proc. 7th Ann. Pacific Climate Workshop, Calif. Dept. of Water Resources, Interagency Ecol. Stud. Prog. Report 26.

Esper, J., Cook, E.R., Peters, K. and Schweingruber, F.H., 2003. Detecting low frequency tree-ring trends by the RCS method. Tree-Ring Research, 59(2), pp.81-98.

Esper, J., Frank, D.C., Wilson, R.J. and Briffa, K.R., 2005. Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophysical Research Letters, 32(7).

Esper, J., Frank, D.C., Timonen, M., Zorita, E., Wilson, R.J., Luterbacher, J., Holzkämper, S., Fischer, N., Wagner, S., Nievergelt, D. and Verstege, A., 2012. Orbital forcing of tree-ring data. Nature Climate Change, 2(12), pp.862-866.

Esper, J., Schneider, L., Smerdon, J.E., Schöne, B.R. and Büntgen, U., 2015. Signals and memory in tree-ring width and density data. Dendrochronologia, 35, pp.62-70.

Frank, D., Esper, J. and Cook, E.R., 2007a. Adjustment for proxy number and coherence in a large-scale temperature reconstruction. Geophysical Research Letters, 34(16).

Frank, D., Büntgen, U., Böhm, R., Maugeri, M. and Esper, J., 2007. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quaternary Science Reviews, 26(25), pp.3298-3310.

Harris, I.P.D.J., Jones, P.D., Osborn, T.J. and Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. International Journal of Climatology, 34(3), pp.623-642.

Luckman, B.H. and Wilson, R.J.S., 2005. Summer temperatures in the Canadian Rockies during the last millennium: a revised record. Climate Dynamics, 24(2-3), pp.131-144.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M. and Francis, R.C., 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the american Meteorological Society, 78(6), pp.1069-1079.

McCarroll, D., Pettigrew, E., Luckman, A., Guibal, F. and Edouard, J.L., 2002. Blue reflectance provides a surrogate for latewood density of high-latitude pine tree rings. Arctic, Antarctic, and Alpine Research, pp.450-453.

Melvin, T.M., Briffa, K.R., Nicolussi, K. and Grabner, M., 2007. Time-varying-response smoothing. Dendrochronologia, 25(1), pp.65-69.

Melvin, T.M. and Briffa, K.R., 2008. A “signal-free” approach to dendroclimatic standardisation. Dendrochronologia, 26(2), pp.71-86.

Melvin, T.M. and Briffa, K.R., 2014. CRUST: Software for the implementation of Regional

Chronology Standardisation: Part 2. Further RCS options and recommendations. Dendrochronologia, 32(4), pp.343-356.

Morice, C.P., Kennedy, J.J., Rayner, N.A. and Jones, P.D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. Journal of Geophysical Research: Atmospheres, 117(D8).

Osborn, T.J., Biffa, K.R. and Jones, P.D., 1997. Adjusting variance for sample-size in tree-ring chronologies and other regional-mean timeseries. Dendrochronologia, 15, pp.89-99.

Rohde, R., Muller, R.A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D. and Wickham, C., 2012. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. Geoinfor Geostat: An Overview, 1(1), pp.1-7.

Rydval, M., Larsson, L.Å., McGlynn, L., Gunnarson, B.E., Loader, N.J., Young, G.H. and Wilson, R., 2014. Blue intensity for dendroclimatology: should we have the blues? Experiments from Scotland. Dendrochronologia, 32(3), pp.191-204.

Rydval, M., Druckenbrod, D., Anchukaitis, K.J. and Wilson, R., 2015. Detection and removal of disturbance trends in tree-ring series for dendroclimatology. Canadian Journal of Forest Research, 46(3), pp.387-401.

Rydval, M., Loader, N.J., Gunnarson, B.E., Druckenbrod, D.L., Linderholm, H.W., Moreton, S.G., Wood, C.V. and Wilson, R., 2017. Reconstructing 800 years of summer temperatures in Scotland from tree rings. Climate Dynamics, pp.1-24.

Schneider, L., Smerdon, J.E., Büntgen, U., Wilson, R.J., Myglan, V.S., Kirdyanov, A.V. and Esper, J., 2015. Revising midlatitude summer temperatures back to AD 600 based on a wood density network. Geophysical Research Letters, 42(11), pp.4556-4562.

Schweingruber, F. 1988. Tree Rings: Basics and Applications of Dendrochronology. Springer, NY.

Solomina, O.N., Bradley, R.S., Jomelli, V., Geirsdottir, A., Kaufman, D.S., Koch, J., McKay, N.P., Masiokas, M., Miller, G., Nesje, A. and Nicolussi, K., 2016. Glacier fluctuations during the past 2000 years. Quaternary Science Reviews, 149, pp.61-90.

Wigley, T.M., Briffa, K.R. and Jones, P.D., 1984. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. Journal of climate and Applied Meteorology, 23(2), pp.201-213.

Wiles, G.C., D'Arrigo, R.D., Villalba, R., Calkin, P.E. and Barclay, D.J., 2004. Century-scale solar variability and Alaskan temperature change over the past millennium. Geophysical Research Letters, 31(15).

Wiles, G.C., D’Arrigo, R.D., Barclay, D., Wilson, R.S., Jarvis, S.K., Vargo, L. and Frank, D., 2014. Surface air temperature variability reconstructed with tree rings for the Gulf of Alaska over the past 1200 years. The Holocene, 24(2), pp.198-208.

Wilson, R.J. and Luckman, B.H., 2003. Dendroclimatic reconstruction of maximum summer temperatures from upper treeline sites in Interior British Columbia, Canada. The Holocene, 13(6), pp.851-861.

Wilson, R., Rao, R., Rydval, M., Wood, C., Larsson, L.Å. and Luckman, B.H., 2014. Blue Intensity for dendroclimatology: The BC blues: A case study from British Columbia, Canada. The Holocene, 24(11), pp.1428-1438.

Wilson, R., K. Anchukaitis, K. Briffa, U. Büntgen, E. Cook, R. D' Arrigo, N. Davi, J. Esper, D. Frank, B. Gunnarson, G. Hegerl, S. Klesse, P. Krusic, H. Linderholm, V. Myglan, Z. Peng, M. Rydval, L. Schneider, A. Schurer, G. Wiles and E. Zorita. 2016. Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. Quaternary Science Reviews, 134, pp.1-18.

Wilson, R., Wilson, D., Rydval, M., Crone, A., Büntgen, U., Clark, S., Ehmer, J., Forbes, E., Fuentes, M., Gunnarson, B.E. and Linderholm, H.W., 2017. Facilitating tree-ring dating of historic conifer timbers using Blue Intensity. Journal of Archaeological Science, 78, pp.99-111.